基于创新扩散理论的学术论文影响力广度研究*

■ 梁国强[1,2] 侯海燕[1] 高桐[3] 孔祥杰[2,3] 胡志刚[1]

[1] 大连理工大学科学学与科学技术管理研究所暨WISE实验室 大连 116023

[2] 印第安纳大学伯明顿分校信息、计算与工程学院 伯明顿 47408 [3] 大连理工大学软件学院 大连 116024

摘要：［目的/意义］采用被引次数衡量学术论文影响力存在诸多弊端。本文认为学术论文影响力包括其传播的深度、速度和广度3个方面，熵可用于衡量学术论文影响力传播的广度。［方法/过程］选择1901－2017年生物医学领域的诺贝尔奖论文作为最具影响力的论文纳入实验组，并根据1∶1配对原则设立对照组，比较两组论文发表后5年内其施引文献所属学科数量、熵以及熵与被引次数的相关性。［结果/结论］实验组中65%以上的论文施引文献学科数量高于对照组；实验组熵的均值介于0.552－0.772，对照组介于0.251－0.481，有显著性差异（$P<0.05$）；两组论文的被引次数与熵的相关性较弱，均小于0.3。结果表明：①70%以上的高影响力论文在发表早期能够影响较多的学科；②采用熵对论文影响力广度进行识别具有可行性。

关键词：创新扩散 熵 诺奖论文 生物医学

分类号：G301

DOI：10.13266/j.issn.0252－3116.2019.02.011

1 引言

创新扩散理论是20世纪60年代美国学者埃弗雷特·罗杰斯（E. M. Rogers）提出的关于解释新思想如何、为何以及以怎样的速度在人群中传播的一种理论[1]。该理论将创新扩散解释为：“以一定的方式随时间在社会系统的各种成员间进行传播的过程，其四要素是：创新、传播渠道、时间和社会系统。”较高的比较优势、兼容性、可尝试性、可见性以及较低的复杂度是促进创新传播最重要的特点。那么何谓扩散？E. M. Rogers认为，扩散就是新思想有意或无意传播的过程。维基百科将扩散解释为分子或原子从高浓度区域向低浓度区域单纯运动直到均匀分布的过程，包含深度、广度和速度3个方面[2]。广度即宽度或横向距离，是衡量宽窄的程度，如图论中著名的广度优先算法，就是从根节点开始，寻找与根节点距离为1的全部节点（此时广度最大），反复迭代直至算法中止[3]；深度即向下或者向里的距离，如深度优先算法，就是从根节点开始，尽可能深地搜索与某节点相连的所有节点，直至所有节点被访问为止[3]；速度则是事物运动的快慢，具有时间属性，物理学中将速度解释为位移对于时间的变化率[4]。

学术网络中，每篇经过同行评审的论文都是新思想、新知识的载体，而论文的引用则记载了这些新思想、新知识的流动[5-7]。有研究认为，论文发表后被引次数的动态变化可视为创新扩散的过程，因此论文的影响力又可细分为深度、广度和速度3个方面[8-10]。本文认为，学术论文影响力的传播速度主要体现在论文发表后达到某一被引量所需的时间。例如，论文A和论文B的被引次数n相同，论文A发表后第3年被引次数累积到n，论文B在发表后第10年累积到n，则论文A影响力传播的速度大于论文B。学术论文影响力传播的深度则反映在论文发表后发生引文级联的次数，即一篇论文发表后被后续论文引用，后续论文又会被后续论文引用，如此往复，从而形成的一个有向无环逐层引用的引文网络[11]。而学术论文影响力传播的广度则体现在论文发表后影响到本领域外其他研究领域的程度[10]。例如，论文C和论文D于同年发表，

* 本文系国家社会科学基金项目“高科技前沿监测中的知识图谱方法与应用研究”（项目编号：14BTQ030）研究成果之一。
作者简介：梁国强（ORCID：0000－0002－9669－4048），博士研究生；侯海燕（ORCID：0000－0002－2790－9973），教授，博士，博士生导师；高桐（ORCID：0000－0002－5464－9433），本科生；孔祥杰（ORCID：0000－0003－2698－3319），副教授，硕士生导师；胡志刚（ORCID：0000－0003－1835－4264），副教授，硕士生导师，通讯作者，E-mail：huzhigang@dlut.edu.cn。
收稿日期：2018－07－01 修回日期：2018－08－12 本文起止页码：91－98 本文责任编辑：易飞

经 n 年后,其被引次数相同,然而论文 C 的施引文献所波及的学科数量远高于论文 D,则论文 C 的影响力广度相对较高。论文影响力的深度、广度和速度是构成学术论文影响力的主要方面,缺一不可。

论文间的引用关系构成了研究期刊间、学科领域间知识流动的基本单元,基于研究人员引用对他们产生影响的论文这一假设,众多学者对科学知识在学科间的传播进行了研究[10,12-16]。例如,T. Van Leeuwen 等[17]认为,学者们往往借助科学出版物来传播他们的学术观点和发现,通过引用关系则可以追踪知识的流动,这些知识通常先在本领域内传播,而后逐渐扩散到其他相邻或相近学科。E. Yan[18]对 Scopus 数据库中的 27 个学科领域的知识创造与扩散过程进行了研究,并借鉴信息熵对学科领域引文活动的多样性进行分析,发现化工、能源与环境科学的影响力增长最快,且大多数学科领域的引文多样性较高,即学科发展较多地受到其他学科的影响。Y. Zhai 等[10]以 Scopus 数据库中收录的 LDA(latent dirichlet allocation)技术的施引文献为数据源,分析了该技术自 2003-2015 年在不同学科领域扩散的过程,发现该技术首先扩散到相邻学科,而后扩散到其他学科领域,并在此过程中,逐渐得到补充、完善和发展。因此,一种学术思想能否以及能在多大程度上在学科间扩散可以反映该思想影响力的广度。

基于创新扩散理论,本文尝试对学术论文影响力的广度进行探索,提出将学术论文施引文献的学科多样性作为衡量学术论文影响力广度的设想。诺贝尔奖(下文简称"诺奖")是科学界的最高奖项,而诺奖论文则是科学界最具影响力和创新性的论文,因此,本研究收集 1901-2017 年生物医学领域获得诺贝尔奖的代表性论文作为实验组,并设立对照组进行比较。文章首先介绍了实验组和对照组数据的获取方法、学术论文施引文献学科多样性的计算方法以及学术论文影响力广度的计算方法;其次,对两组论文施引文献的学科多样性、影响力广度、熵与被引次数的相关性及采用熵衡量论文影响力广度的可行性进行了分析;最后,对本研究进行讨论和总结。

2 数据来源与方法

2.1 数据来源

本研究的数据全部源于 Web of Science(WoS)数据库,收集生物医学领域的高被引诺奖论文作为实验组,并按 1:1 比例,随机选取该领域非诺奖论文作为对照组,如图 1 所示:

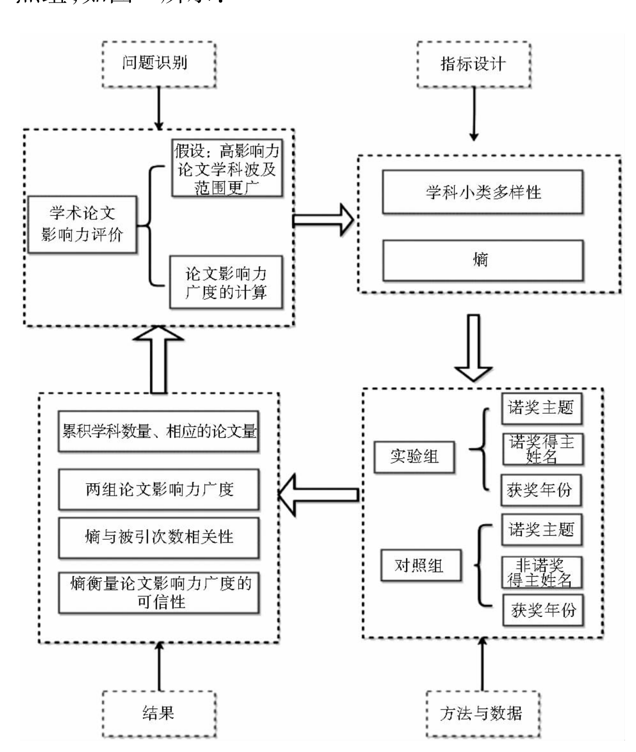


图 1 学术论文影响力广度研究框架

1901-2017 年,生物医学领域有 214 位研究人员被授予诺贝尔奖。依据诺奖委员会提供的获奖得主名单、获奖时间以及颁奖词,本研究提取了诺奖得主的姓名、获奖时间、获奖主题,并将诺奖得主在授予诺奖前发表的与获奖主题相关,且被引次数最高的两篇论文纳入实验组。例如:根据诺奖委员会官方网站显示,H. Dam 因发现维生素 K 而获得 1943 年的诺贝尔生理医学奖,因此本研究将"Vitamin K"作为检索主题词,检索 H. Dam 在 1943 年以前发表的全部期刊论文中被引次数最高的 2 篇。但并非全部诺奖得主的论文都和 H. Dam 的一样比较容易确定检索主题词,例如诺奖委员会官网显示,卡尔·兰德施泰纳(K. Landsteiner)因发现人类的 ABO 血型系统而获得 1930 年的诺贝尔奖,其获奖主题词并不明确,因此本研究获取了维基百科上提供的关于 K. Landsteiner 的个人信息,将"血液""抗原""血清""血清学"作为 K. Landsteiner 的获奖关键词,检索他在 1930 年以前发表的全部期刊论文中被引次数最高的两篇。最终,本研究获取了 389 篇诺奖论文及发表后 5 年内的 67 724 篇施引文献,纳入实验组。

对照组属于 1:1 配对设计的随机对照组,即选择截至诺奖得主获奖前,与该得主研究主题相同但未获得诺奖的研究人员发表的论文。例如,对 H. Dam 的

诺奖论文选取对照时,本研究将"Vitamin K"作为检索主题词,在 WoS 数据库中检索 1943 年以前发表的除 H. Dam 以外的研究人员的期刊论文,随机抽取两篇作为对照。最终获取了 389 篇非诺奖论文及发表后 5 年内的 17 353 篇施引文献。

2.2 研究方法

2.2.1 论文施引文献学科多样性的计算　依据科睿唯安(Clarivate Analytics)的学科划分标准,将科学领域分为艺术与人文(Arts & Humanities)、医学领域(Clinical, Pre - Clinical & Health)、工程与技术(Engineering & Technology)、生命科学(Life Sciences)、物理科学(Physical Sciences)和社会科学(Social Sciences)六大学科门类,每个学科门类下,包含不同的学科细分领域(学科小类)共 256 个[19]。学科多样性算法的基本思想是计算论文的施引文献所属学科领域中不属于该论文所属领域的数量,具体步骤为:将所有医学领域的细分领域积分设为 0,所有非医学领域的细分领域积分设为 1,记录两组数据中每篇论文发表后 5 年内的得分,计算方法如式(1)所示:

$$S_i = \sum_{j=1}^{n} x_{ij},\ x_{ij} = \begin{cases} 1, & WC_{ij} \in D - D_{medicine} \\ 0, & WC_{ij} \in D_{medicine} \end{cases} \qquad 式(1)$$

其中,S_i 为论文 i 的累积学科数量,n 是论文 i 所波及学科的数量,x_{ij} 为论文 i 波及的第 j 个学科小类的数量,若该学科属于非医学领域($D - D_{medicine}$)则认为其波及的学科数量为 1,反之数量为 0。在计算中,若某论文的施引文献所属学科分类相同,不会重复计算。同时,由于医学大类中包含 47 个小类,因此一篇论文的累积积分理论上不应超过 209(即 256 个学科减去 47 个医学领域的学科)。

2.2.2 论文影响力广度的计算　熵最初是根据是热力学第二定律引申出来的概念,是对一个系统能量衰竭程度即混乱度的度量单位,系统越混乱,熵越高。1948 年,香农(Shannon)将熵引入信息论中,认为熵是一个随机事件的不确定性或信息量的度量,不确定性越高,熵越高,当随机变量概率相同时,熵达到最大。信息熵的提出大大促进了信息论的发展,奠定了现代信息论的理论基础。在学术网络中,学科的多样性可以用熵来衡量[20],本文将论文 j 的施引文献所属学科领域中各学科所占的比例作为条件概率 p(j),当论文 j 只属于 GIPP 学科分类标准中的某一个学科细分领域时,论文 j 的熵为该论文施引文献所属各细分学科领域在该论文施引文献所属学科总量中所占的比率。当论文 j 属于 2 个及以上学科时,论文 j 的熵为论文 j 所属的各学科细分领域与论文 j 的施引文献学科细分领域间的熵的均值作为论文 j 的熵。当论文 j 的施引文献学科细分领域小于等于 1 时,此时熵为 0,理论上,论文 j 施引文献所属的学科多样性越高,熵值越大。根据熵的定义,本文熵的计算公式如式(2)所示:

$$H(j) = -\sum_{j}^{n} p_j * \lg(p_j) \qquad 式(2)$$

其中,$H(j)$ 是论文 j 的熵,p_j 是条件概率,即论文 j 的施引文献所属学科细分领域与该论文施引文献所属学科细分领域之和的比率。

3 结果

3.1 两组论文施引文献所属学科分布

3.1.1 累积学科分布　本部分主要从宏观上了解两组论文发表后 5 年内历年累积的学科数量。如图 2 所示,两组论文的学科累积数量逐年增加,实验组论文的学科累计数量高于对照组。

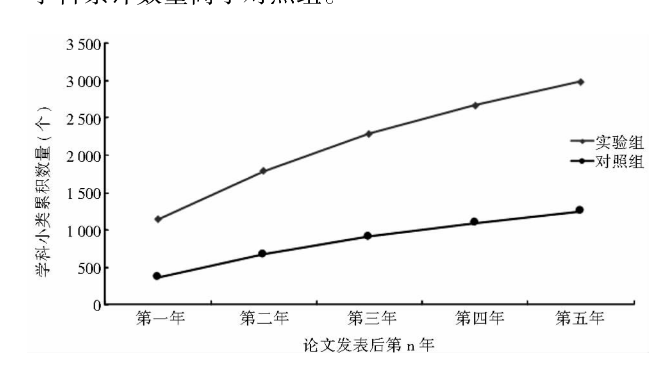


图 2　两组论文发表后 5 年内学科累积数量

3.1.2 学科数量与论文量分布　为进一步了解两组数据中,不同学科数量下对应的论文数量,该部分对两组数据涉及的学科数量与对应论文量进行了统计。如图 3 所示,实验组波及 5 个以上学科数量的论文高于对照组。进一步分析发现,实验组论文发表后第一年至第五年,65% 以上的论文施引文献学科波及数量高于对照组,即论文发表后第一年至第五年,实验组中分别含 253 篇(占 65.04%)、251 篇(占 64.52%)、261 篇(占 67.10%)、265 篇(占 68.12%)和 267 篇(占 68.64%)论文的施引文献学科数量高于对照组。

3.2 两组论文影响力的广度

本文将学术论文发表后其施引文献所属学科的多样性作为衡量论文影响力广度的方法,并将熵应用于学科多样性的测算。如图 4 所示,两组论文的熵值自论文发表后第一年至第五年的均值逐年增加,实验组的熵的均值集中在 0.552 - 0.772 之间,对照组则集中在 0.251 - 0.481 之间,经 U 检验, 实验组论文的熵值

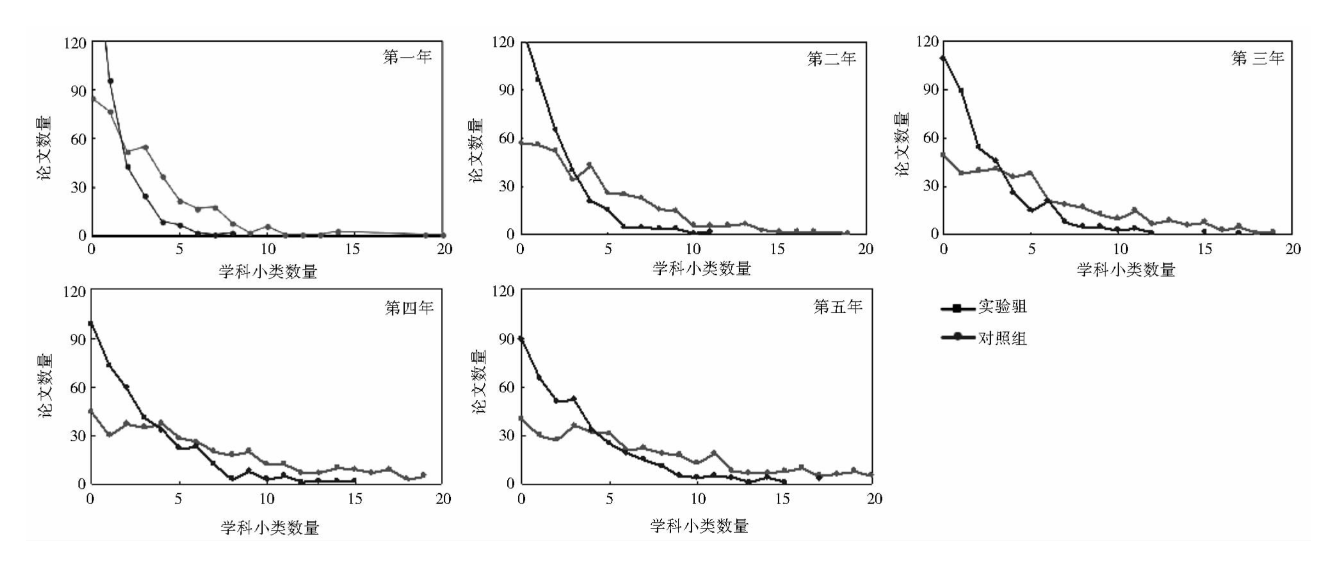


图 3 两组论文学科数量与论文量

第一年
均数 =0.552
标准差 =0.347

第二年
均数 =0.666
标准差 =0.351

第三年
均数 =0.719
标准差 =0.350

第四年
均数 =0.754
标准差 =0.349

第五年
均数 =0.772
标准差 =0.349

A

熵 (H)

论文编号

第一年
均数 =0.251
标准差 =0.289

第一年
均数 =0.381
标准差 =0.330

第三年
均数 =0.426
标准差 =0.351

第四年
均数 =0.455
标准差 =0.358

第五年
均数 =0.481
标准差 =0.367

B

熵 (H)

论文编号

图 4 实验组(图 A)和对照组(图 B)的熵值

大于对照组($P < 0.05$)。进一步调查发现,论文发表后第一年至第五年,实验组中分别有273篇(占70.36%)、268篇(占69.07%)、274篇(占70.62%)、276篇(占71.13%)、276篇(占71.13%)论文的熵值高于对照组。

3.3 两组论文被引次数与熵的相关性

图5为实验组与对照组两组被引次数差异的年份变化示意图,其显示实验组被引次数高于对照组的论文自发表后第一年至第五年,呈线性增长的趋势($R^2 = 0.976$),实验组被引次数小于对照组的论文数量围绕76篇(19.54%)上下波动,而实验组与对照组被引次数相等的论文数量自发表后第一年至第五年,呈线性递减的趋势($R^2 = 0.828$)。

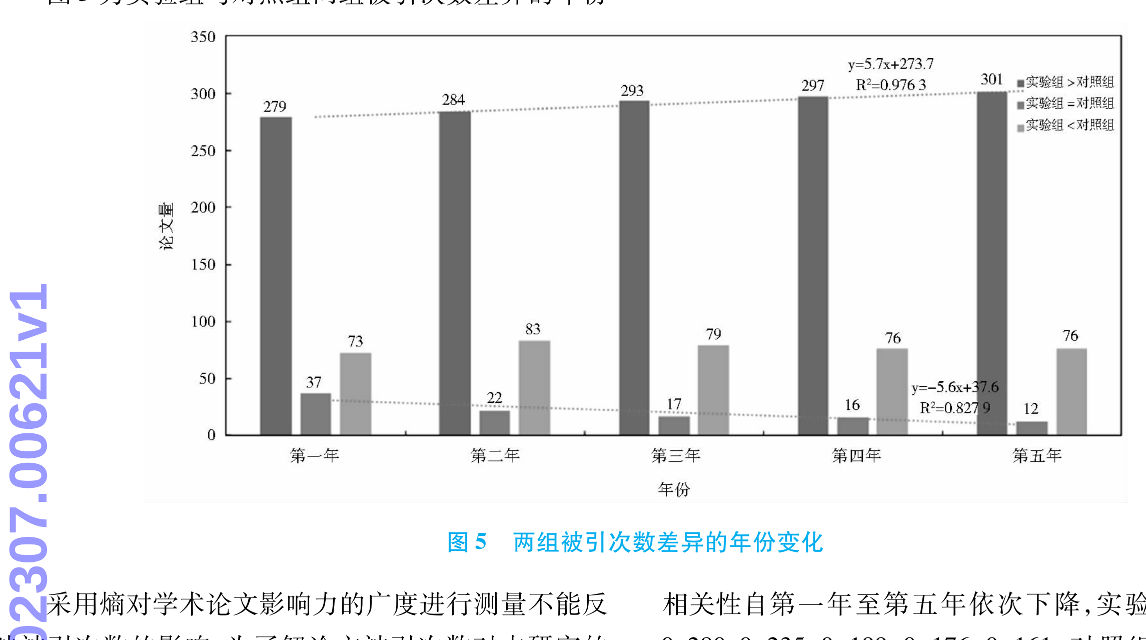


图5 两组被引次数差异的年份变化

采用熵对学术论文影响力的广度进行测量不能反映被引次数的影响,为了解论文被引次数对本研究的潜在影响,本部分对两组数据中熵与被引次数的相关性进行了研究,结果发现两组论文的被引次数与熵的相关性自第一年至第五年依次下降,实验组分别为:0.280,0.235,0.199,0.176,0.161;对照组为:0.240,0.143,0.109,0.094,0.078,相关性较弱。具体如图6所示:

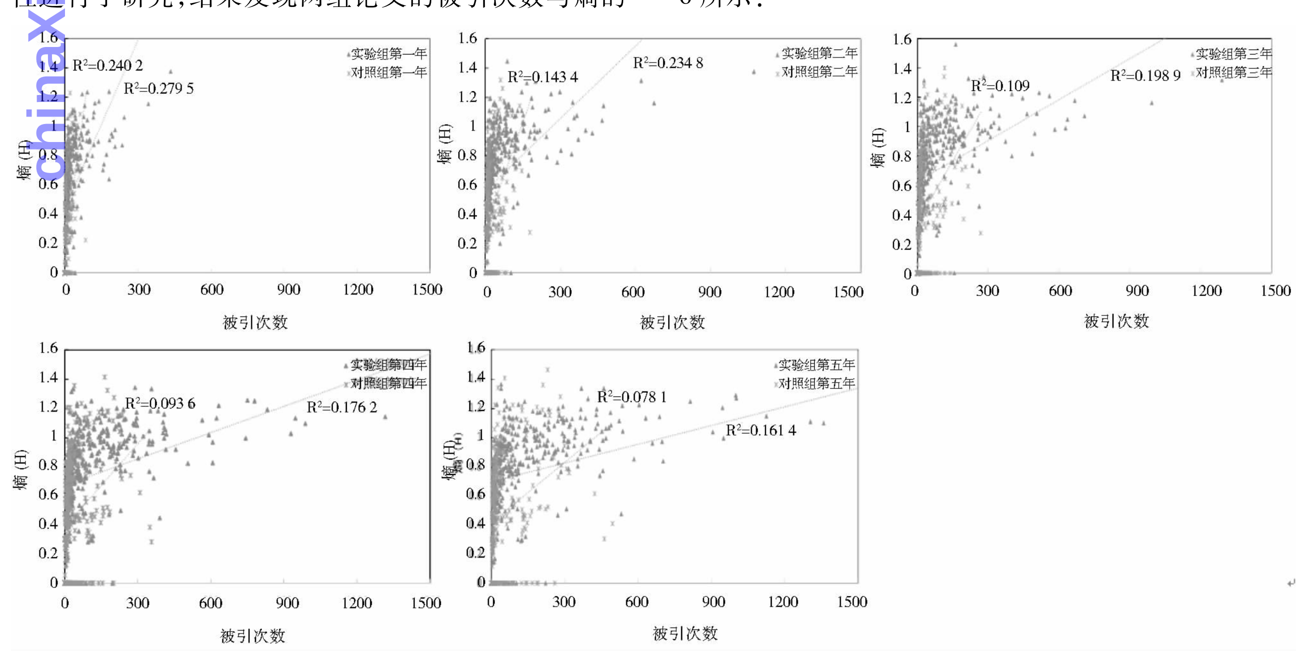


图6 实验组与对照组论文熵与被引次数的相关性

3.4 采用熵衡量学术论文影响力广度的可行性

假设实验组论文的熵大于对照组,则相应的该组论文的施引文献所属学科的多样性要高于对照组,反之亦成立。因此,本部分将两组数据熵的差异与

学科数量差异的年代变化进行了统计。结果显示，论文发表第一年至第五年，两组数据中熵的差异与学科细分领域数量的差异的变动一致的比例分别为77.84%、55.67%、75.26%、75%、、73.97%，如表 1 所示：

表 1 两组论文的熵值与学科细分领域比较

	熵(H)	学科细分领域数量(S)		
		实验组 > 对照组	实验组 = 对照组	实验组 < 对照组
第一年	实验组 > 对照组	232	31	12
	实验组 = 对照组	6	32	6
	实验组 < 对照组	15	17	38
第二年	实验组 > 对照组	187	44	37
	实验组 = 对照组	7	14	4
	实验组 < 对照组	58	22	15
第三年	实验组 > 对照组	229	20	25
	实验组 = 对照组	0	15	4
	实验组 < 对照组	31	16	48
第四年	实验组 > 对照组	229	22	25
	实验组 = 对照组	0	15	2
	实验组 < 对照组	35	13	47
第五年	实验组 > 对照组	228	20	28
	实验组 = 对照组	0	13	1
	实验组 < 对照组	38	14	46

4 讨论

在创新驱动发展的背景下，对学术论文的影响力做出公正、客观的评价，有利于管理者和政策制定者及时掌握前沿性成果，让有限的资源最大限度地推动科学进步，同时，也关系到科研人员职称评定、薪酬设置、福利以及是否获得资助，影响其工作热情。

早在 1927 年，美国波莫纳学院(Pomona college)化学院的 P. L. Gross 等[21]等在讨论大学图书馆建设及化学学科教育问题时，就开始采用被引次数对科研工作的重要性进行评估。自此以后，引文分析普遍应用于国家科学政策制定和学科发展、研究团体、期刊、个人及论文的评价[22-24]。如：经济合作与发展组织(OECD)在报告中采用被引次数排名在 top10% 的论文来评价不同国家的科研表现，发现被引次数世界排名 top10% 的论文中，我国从 2005 年的不足 4% 上升到 2016 年的 14%，科研表现仅次于美国[25]。自 2013 年始，科睿唯安(Clarivate Analytics)公司每年发布的《研究前沿》报告也采用了被引次数排名在 top1% 的论文来捕捉自然科学和社会科学领域主要的热点前沿、新兴前沿等。但是，学术网络是一个动态、自组织、复杂的网络系统，采用论文的累积被引次数来衡量其学术影响力不利于新发表论文的评价而且受"马太效应"的影响[5]，另外，研发发现在研究人员的学术生涯中，被引次数最高的论文不一定是该研究人员水平最高的论文[26]。

为弥补被引次数的不足，研究人员进行了大量探讨，从 h 指数[27]到 g 指数[28]，从影响因子[29]到特征因子(eigenfactor)[30]等基于被引次数的评价指标应运而生。这些指标逐渐完善了单纯依赖被引次数进行学术影响力评价的弊端，但仍有一定的不足。例如，采用 h 指数和 g 指数虽兼顾了论文的数量和被引量，但采用该指标不适合对年轻科学家进行评价。影响因子的设计初衷是用于期刊筛选，而如今已广泛用于学术论文质量的评价，但该指标受文章类型、研究领域等因素的影响，且发表在同一刊物上的论文在经过一段时间后其被引次数差异较大。特征因子则借鉴了 PageRank 算法的一些思想，认为高影响力期刊引用的期刊的影响力也高，更加注重了期刊的质量，并衍生出论文影响分值(article influence score)这一指标，但对相对影响力较低的期刊来讲，特征因子的区分度不够，且计算方式复杂[31]。

基于创新扩散理论，本文提出学术论文的影响力分为深度、速度和广度 3 个维度，认为可以通过计算论文发表后施引文献的学科多样性来反映论文影响力的广度。文章以生物医学领域的诺奖论文为例，将这些论文视为高影响力论文，并按 1∶1 配对原则收集该领域的非诺奖论文作为对照。考虑到论文发表年份不一致，在时间窗口设置时，本研究对两组论文发表后第 1 年到第 5 年的施引文献学科多样性进行比较，以增强数据的可比性。

结果发现，论文发表后第一年至第五年，实验组 65% 以上的论文的施引文献所属学科数量高于对照组，广泛分布于 5 个及以上的学科中。而对两组论文相应时间段内的被引次数的统计显示，实验组论文中 70% 以上的论文被引次数高于对照组，提示被引次数高的论文其施引文献的学科数量相对较高。而熵是衡量系统复杂性、混乱度的重要指标，采用熵对施引文献学科的测算则可以减少或避免被引次数的潜在影响，论文发表后 5 年内，实验组熵值的均值介于 0.552 - 0.772，对照组介于 0.251 - 0.481，有显著性差异($p < 0.05$)，且两组论文的熵与被引次数的相关性较低，表明实验组论文的影响力广度高于对照组。本研究显示，部分高影响力论文在最初发表的 5 年内，其被引次数小于或等于对照组的论文的比例分别为 28.28%，

26.99%,24.68%,23.65%,22.62%,其中,实验组被引次数小于对照组的论文数量围绕76篇(19.54%)上下波动,提示论文发表早期,高影响力论文的被引次数不一定高于普通论文,依据论文被引次数对论文影响力早期识别存在一定的偏差。

5 结论

本研究得出如下结论:①70%以上的高影响力论文在发表早期能够影响较多的学科;②采用熵对论文影响力广度进行识别具有可行性。

在学术网络中,一些论文并没有产生广泛的影响力,一方面与该论文本身创新性不足有关,另一方面如果该论文创新性、复杂度过高,从而超出人们的常识、理解能力时,也不容易扩散,而成为"睡美人"论文。当学术论文在创新与保持传统之间保持了必要的张力时,往往能收获最大的影响力[6, 32-33]。

本研究处于探索阶段,存在一定的局限性:①熵作为衡量学科多样性的指标,其本身无法反映论文被引次数的大小。例如,论文A和B都源于医学领域,其施引文献的所属学科均来自艺术、医学、哲学3个学科,论文A的被引次数为100,论文B为10,但二者的熵相同。②本研究对学科多样性的研究默认各学科领域的相似性相同,但实际上各学科间相似性存在差异,例如泌尿科学与神经科学之间的相似性较泌尿科学与语言学的相似性高,当论文A的施引文献所属学科类型为泌尿科学与神经科学,论文B为泌尿科学与语言学时,二者的熵相同。今后,将把学科间相似性考虑到指标的设计中,从而更加客观地比较论文学术影响力的广度。而且,由于单一的指标无法全面评估一篇论文的影响力,今后会研究如何将学术论文影响力的3个维度综合考量,对学术论文的影响力进行早期识别。

致谢:感谢Filipi Nascimento Silva对学科多样性的相关算法提供的帮助,感谢国家留学基金委的资助。

参考文献:

[1] ROGERS E M. Diffusion of innovations [M]. New York: The Free Press, 2010.

[2] ZHANG L, PENG T－Q. Breadth, depth, and speed: diffusion of advertising messages on microblogging sites [J]. Internet research, 2015, 25(3): 453－470.

[3] 伊斯利,克莱因伯格. 网络、群体与市场 [M]. 北京:清华大学出版社, 2011.

[4] JONES A Z. What is velocity in physics? [EB/OL]. [2018－10－08]. https://www.thoughtco.com/velocity－definition－in－physics－2699021.

[5] WANG D, SONG C, BARABASI A L. Quantifying long－term scientific impact [J]. Science, 2013, 342(6154): 127－132.

[6] UZZI B, MUKHERJEE S, STRINGER M, et al. A typical combinations and scientific impact [J]. Science, 2013, 342(6157): 468－472.

[7] GARFIELD E. Citation indexes for science: a new dimension in documentation through association of ideas [J]. Science, 1955, 122(3159): 108－111.

[8] MIN C, SUN J, DING Y. Quantifying the evolution of citation cascades [J]. Proceedings of the Association for Information Science and Technology, 2017, 54(1): 761－763.

[9] MIN C, DING Y, LI J, et al. Innovation or imitation: the diffusion of citations [J]. Journal of the Association for Information Science and Technology, 2018, 69(10):1271－1282.

[10] ZHAI Y, DING Y, WANG F. Measuring the diffusion of an innovation: a citation analysis [J]. Journal of the Association for Information Science and Technology, 2018, 69(3): 368－379.

[11] 闵超, DING Y, 李江, 等. 单篇论著的引文扩散 [J]. 情报学报, 2018, 37(4): 341－350.

[12] KISS I Z, BROOM M, CRAZE P G, et al. Can epidemic models describe the diffusion of topics across disciplines? [J]. Journal of informetrics, 2010, 4(1): 74－82.

[13] CHEN C, HICKS D. Tracing knowledge diffusion [J]. Scientometrics, 2004, 59(2): 199－211.

[14] JAFFE A B, TRAJTENBERG M. Flows of knowledge from universities and federal laboratories: modeling the flow of patent citations over time and across institutional and geographic boundaries [J]. Proceedings of the National Academy of Sciences, 1996, 93(23): 12671－12677.

[15] TIJSSEN R J W. Global and domestic utilization of industrial relevant science: patent citation analysis of science-technology interactions and knowledge flows [J]. Research policy, 2001, 30(1): 35－54.

[16] ZHU Y, YAN E. Dynamic subfield analysis of disciplines: an examination of the trading impact and knowledge diffusion patterns of computer science [J]. Scientometrics, 2015, 104(1): 335－359.

[17] VAN LEEUWEN T, TIJSSEN R. Interdisciplinary dynamics of modern science: analysis of cross-disciplinary citation flows [J]. Research evaluation, 2000, 9(3): 183－187.

[18] YAN E. Disciplinary knowledge production and diffusion in science [J]. Journal of the Association for Information Science and Technology, 2016, 67(9): 2223－2245.

[19] SILVA F N, RODRIGUES F A, OLIVEIRA O N, et al. Quantifying the interdisciplinarity of scientific journals and fields [J]. Journal of informetrics, 2013, 7(2): 469－477.

[20] Clarivate Analytics. GIPP mapping table [EB/OL]. [2018－10－08]. http://ipscience－help.thomsonreuters.com/inCites2Live/indicatorsGroup/aboutHandbook/appendix/mappingTable.html.

[21] GROSS P L, GROSS E M. College libraries and chemical education [J]. Science, 1927, 66(1713): 385－389.
[22] LU C, DING Y, ZHANG C. Understanding the impact change of a highly cited article: a content-based citation analysis [J]. Scientometrics, 2017, 112(2): 927－945.
[23] ZENG A, SHEN Z, ZHOU J, et al. The science of science: from the perspective of complex systems [J]. Physics reports, 2017, 714－715: 1－73.
[24] 梁国强,侯海燕,任佩丽,等. 高质量论文使用次数与被引次数相关性的特征分析 [J]. 情报杂志, 2018, 37(4): 147－153.
[25] OECD. OECD science, technology and industry scoreboard 2017: the digital transformation [EB/OL]. [2018－10－08]. https://read.oecd－ilibrary.org/science－and－technology/oecd－science－technology－and－industry－scoreboard－2017_9789264268821－en#page1.
[26] IOANNIDIS J P, BOYACK K W, SMALL H, et al. Bibliometrics: is your most cited work your best? [J]. Nature, 2014, 514 (7524): 561－562.
[27] HIRSCH J E. An index to quantify an individual's scientific research output [J]. Proceedings of the National Academy of Sciences of the United States of American, 2005, 102(46): 16569－16572.
[28] EGGHE L. Theory and practise of the g－index [J]. Scientometrics, 2013, 69(1): 131－152.
[29] GARFIELD E. The history and meaning of the journal impact factor [J]. Journal of the American Medical Association, 2006, 295 (1): 90－93.
[30] BERGSTROM C. Eigenfactor: measuring the value and prestige of scholarly journals [J]. College & research libraries news, 2007, 68(5): 314－316.
[31] 任胜利. 基于引证网络分析期刊和论文的重要性 [J]. 中国科技期刊研究, 2009, 20(3): 415－418.
[32] FOSTER J G, RZHETSKY A, EVANS J A. Tradition and innovation in scientists' research strategies [J]. American sociological review, 2015, 80(5): 875－908.
[33] WANG J, VEUGELERS R, STEPHAN P. Bias against novelty in science: a cautionary tale for users of bibliometric indicators [J]. Research policy, 2017, 46(8): 1416－1436.

作者贡献说明:
梁国强:提出研究设想,设计研究框架,收集、分析数据,撰写论文;
侯海燕:负责论文指导、修改;
高桐:负责收集、分析数据;
孔祥杰:参与讨论,完善研究框架;
胡志刚:负责进行论文指导。

Impact Breadth of Scientific Papers Based on Innovation Diffusion Theory

Liang Guoqiang[1,2] Hou Haiyan[1] Gao Tong[3] Kong Xiangjie[2,3] Hu Zhigang[1]
[1] WISE Lab of Dalian University of Technology, Dalian 116023
[2] School of Informatic and computing, Indiana University Bloomington, Bloomington 47408
[3] School of Software, Dalian University of Technology, Dalian 116024

Abstract: [**Purpose/significance**] Using citations as a measurement of the impact of an article have long been criticized. This study argues that the impact of scientific paper includes its depth, speed, and breadth in diffusion. We can measure the impact breadth of an article by using entropy. [**Method/process**] This study regards Nobel Prizes winning articles in Physiology and Medicine field as the most influential scientific research, collecting Nobel Prizes winning articles into the test group and matching them at a ratio of 1:1 into the control group. The citing articles' disciplinary diversity, within five years after these papers' publishing, was explored. In addition, this study employs entropy to measure the diversity. [**Result/conclusion**] 65 percent of articles in the test group have relatively higher disciplinary diversity compared to the control group. The values of entropy in the test group are between 0.552 and 0.772, and between 0.251 and 0.481 in the control group ($p<0.05$). There is a weak correlation, which below 0.3, between citation counts and values of entropy. The paper argues that above 70 percent of highly influential articles have a high disciplinary diversity in their early stage, and it is possible by using entropy as an indicator to measure the breadth of an article's impact.

Keywords: innovation diffusion entropy Nobel Prizes winnings articles Physiology and Medicine